Modeling Individual Claims with the Generalized Linear Transformed Mixture Cluster-Weighted Model

Nik Počuĉa

Supervisor: Dr. Paul McNicholas

Department of Mathematics and Statistics

McMaster University

November 29, 2017

Acknowledgements

Abstract

In this paper, we extend the assumption behind the generalized linear mixture cluster-weighted models (CWM) to accommodate non-Gaussian distribution of the continues covariates which is more appropriate for the insurance applications. Generalized linear mixture cluster-weighted models are considered as a flexible family of mixture models for fitting the joint distribution of a random vector composed of a response variable and a set of continues and discrete covariates. However, these models assume that the continuous covariates come from the Gaussian distribution. If this assumption is violated, then the current CWM may provide suboptimal results. We developed a generalized linear transformed mixture cluster-weighted model (TCWM) that considers a transformed distribution of covariates. A simulation study showed that the TCWM performs well for different settings. A real data set based on French automobile policies is used to illustrate the application of the proposed method.

Contents

Chapter 1

Introduction

Predictive modelling gained a lot of attention in the past decade in the area of actuarial science, risk management, and insurance in general. While the term predictive modelling has been used in many other areas, in context of insurance, it is referred to as a process of using statistics in estimating the insurance cost (**?**). Various predictive models are used in the area of actuarial science with generalized linear models (GLMs) being the most popular tools actively integrated in pricing, reserving, and underwriting of property and casualty insurance. A generalized linear model is a framework for extending the concept of a linear regression. A GLM consists of three components. The conditional distribution of Y | X must belong to the exponential family, A link function g that relates the conditional mean and the covariates must be defined, and finally a linear predictor denoted as $\eta = X\beta$ for design matrix X and vector of coefficients β. (**?**).

The origin of GLMs began with the paper of Nelder and Wedderburn (1972). The text **?** explained the framework in detail, and then was further expanded with the review of applications by **?**. GLMs are flexible tools for handling non-normal data, typically observed in frequency and severity of claims. GLMs can also handle additive, multiplicative, or logistic effect of explanatory variables on a transformation of the mean. Several extensions of GLMs with applications in actuarial science have been developed by **?**, **?**, **?**, and **?**. GLMs with categorical response variables are usually used for modelling frequency while GLMs with continues response variables can be used for modelling severity of claims. The results from both of these analyses are used in pricing. The concept of a mixture of GLMs is defined as a partitioned probability space where each component of the mixture is a separate GLM (**?**). Consider a mixture model with K components of the form:

$$f(y|x, \phi) = \sum_{k=1}^{K} \pi_k f(y|x, \theta_k)$$

In which π_k is the prior probability of component k, and θ_k is the component parameter. $\phi = (\pi_1 \ldots \pi_k, \theta_1' \ldots \theta_k')'$. Mixture of GLMs with applications outside of actuarial science was considered by (**?**). Other applications of mixtures of GLMs are explored in other fields such as market segmentation, see **?**, in biology or medicine, see **?**, **?** and **?**. Several actuarial applications of GLMs are based on gamma and log-normal regressions discussed by Boucher and Charpentier (Chapter 14) in the book by **?**. It is observed that the coefficients of these models will be impacted by claim size (i.e., the outliers may effect gamma regression more than log-normal due to the logarithmic scale). To overcome this issue, the authors proposed modelling large and normal size claims separately with a predetermined threshold. The probability of a large or standard claim can be modelled with separate

logistic regressions. The extension of this model is considered as a multinomial regression where the mixing probabilities associated with each category of claim size are based on a multinomial random variable. This is an example of a mixture model where the population of claims is manually grouped based on a similar loss experience within a group.

Rather than using manual thresholds to classify claims into groups, we propose a statistical method based on a finite mixture of GLMs. This approach would also allow for finding an optimal number of groups based on a given clustering criteria (e.g. BIC). The estimated regression coefficients will vary for each GLM component. We introduce the concept of a generalized linear mixture cluster-weighted model (CWM) proposed by **?**. CWM is a flexible family of mixture models for fitting the joint distribution of a random vector. This random vector is composed of a response variable and a set of mixed-type covariates with the assumption that continues covariates come from Gaussian distribution. Like the mixture model, CWM is a partitioned mixture of a weighted model in the form:

$$f(\boldsymbol{x}, \boldsymbol{y}; \Phi) = \sum_{g=1}^{G} \tau_j f(\boldsymbol{y}|\boldsymbol{x}, \vartheta_j) f(\boldsymbol{v}, \theta_j') f(\boldsymbol{w}, \theta_j'')$$

We have that $x = (v, w)$ where x is decomposed into two separate variables. v denotes the continuous variables, and w denotes the discrete variables. The $f(x, y, \Phi)$ term denotes a joint distribution of x, y with parameters Φ. The joint distribution is partitioned into G groups, with prior probability component τ_j. The $f(\boldsymbol{y}|\boldsymbol{x}, \vartheta_j)$ is the conditional density of y given x, with parameters ϑ_j. $f(\boldsymbol{v}, \theta_j')$ is the marginal density of the continuous covariates v with parameters ϑ_j'. Finally the $f(\boldsymbol{w}, \theta_j''$ is the marginal density of discrete covariates with parameters ϑ_j''. Additional applications of CWM's can be found in the papers written by **?**, **?**, and **?** in a context of media technology. Some extensions of this class of models have been considered by **?**, **?**, and **?**.

In this paper, we propose an extension of CWM to allow covariates to belong to a non-Gaussian family. Our method requires a transformation to be applied to the non-Gaussian covariates, we define our proposed model as generalized linear transformed cluster-weighted model (TCWM). The justification for this approach is that some insurance rating variables may not follow a Gaussian distribution on a positive domain (e.g. car exposure or population density).

Evidence suggests that insurance losses exhibit multimodality. Mixture models of univariate insurance data based on non-Gaussian distributions have been developed by **?**. **?** and **?** proposed mixture models of Erlang distributions. **?** worked toward developing a unified approach in fitting loss distributions. They determined that finite mixture models, using a small number of base distributions, provided an ideal set. Mixtures were also studied in a regression setting. (**?**) developed a mixture of bivariate Poisson regression that uses zero-inflated bivariate Poisson regression as a special case. The authors fitted the model using the EM algorithm. Another type of Poisson mixture model for count data was proposed by **?** and applied to a group life Insurance portfolio.

Motivated by the idea of mixture modeling, we extend the idea of univariate mixture modeling to the mixture modeling of regressions including the GLMs, where losses are modeled as a function of several covariates. This approach would allow for finding the groups of observations with similar regression coefficients that can be further considered in the general insurance pricing.

This paper is organized as follows. Section 2 presents the proposed TCWM model for mixture of GLMs. Section 3 shows an application of the TCWM model. Section 4 describes the simulation study. Section 5 discuss the results obtain from a real data set when modelling individual insurance claims. Conclusion is provided in Section 6.

1.1 Methodology

The General Linear Transformed Cluster-Weighted Model

In this section, we present TCWM approach as the extension of CWM approach for modeling amount or claims count, $\boldsymbol{Y}$, as a function of several covariates $\boldsymbol{x}$. The CWM model was proposed by **?** as a class of mixture of regression models with random covariates and is currently implemented in CRAN library as flexCWM.

Suppose that $\boldsymbol{Y} = (Y_1, \dots, Y_N)$ is a vector of independent random variables with the density function of a distribution from the exponential family. For this family of distributions, $\boldsymbol{Y}$ is related to a set of variables $\boldsymbol{x} = (x_1, \dots, x_p)$ through the following linear relationship

$$\eta = g[E(\boldsymbol{y}|\boldsymbol{x})] = \boldsymbol{X}\boldsymbol{\beta}, \tag{1.1.1}$$

where $\boldsymbol{\beta}$ is $p \times 1$ vector of parameters, $\boldsymbol{X}$ is an $N \times p$ design matrix, η is the linear predictor, and $g[E(\boldsymbol{y}|\boldsymbol{x})]$ is a link function in which it is considered as a simple mathematical function in the original formulation of GLMs by **?**. In this model, $\boldsymbol{Y}$ is referred to as a response variable and $\boldsymbol{x}$ is a vector of continuous and discrete type explanatory variables (i.e., covariates).

Let $(\boldsymbol{X}', \boldsymbol{Y})'$ be a vector defined on some space $\mathcal{R}$ with values in $\mathcal{X} \times \mathcal{Y}$. If there exist G partitions of $\mathcal{R}$, defined as $\mathcal{R}_1, \dots, \mathcal{R}_G$, **?** defined the CWM as a finite mixture of GLMs where the joint distribution $f(\boldsymbol{x}, \boldsymbol{y})$ of $(\boldsymbol{X}', \boldsymbol{Y})'$ is expressed as follows

$$f(\boldsymbol{x}, \boldsymbol{y}) = \sum_{g=1}^{G} \tau_g f(\boldsymbol{y}|\boldsymbol{x}, \mathcal{R}_g) f(\boldsymbol{x}, \mathcal{R}_g) \tag{1.1.2}$$

where $f(\boldsymbol{y}|\boldsymbol{x}, \mathcal{R}_g)$ and $f(\boldsymbol{x}, \mathcal{R}_g)$ are conditional and marginal distributions of $(\boldsymbol{X}', \boldsymbol{Y})'$ and τ_g. Further, the weights $\{\tau_1, \dots, \tau_G\}$, with $\sum_{g=1}^{G} \tau_g = 1$ and $\tau_g > 0$, are defined as the unknown proportions of observations in each gth group.

? proposed mixture models for fitting the joint distribution of a random vector $(\boldsymbol{X}', \boldsymbol{Y})'$ by splitting the covariates into continuous and discrete as $\boldsymbol{X} = (\boldsymbol{V}', \boldsymbol{W}')'$ so that their corresponding marginal distributions can be multiplied. Therefore, the model in Equation 1.1.2 can be written as

$$f(\boldsymbol{x}, y; \Phi) = \sum_{g=1}^{G} \tau_g f(y|\boldsymbol{x}, \vartheta_g) f(\boldsymbol{x}, \theta_g) = \sum_{g=1}^{G} \tau_g f(y|\boldsymbol{x}, \vartheta_g) f(\boldsymbol{v}, \theta_g') f(\boldsymbol{w}, \theta_g'') \tag{1.1.3}$$

where $\boldsymbol{v}$ and $\boldsymbol{w}$ are the vectors of continues and discrete covariates respectively, $f(y|\boldsymbol{x}, \vartheta_g)$ is a conditional density of $y|\boldsymbol{x}$ with parameter ϑ_g, $f(\boldsymbol{v}, \theta_g')$ is the marginal distribution of $\boldsymbol{v}$ with parameter θ_g', $f(\boldsymbol{w}, \theta_g'')$ is the marginal distribution of $\boldsymbol{w}$ with parameter θ_g'', and $\Phi = (\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\vartheta})$ includes all model parameters. Conditional distribution $f(y|\boldsymbol{x}, \vartheta_g)$ is assumed to belong to the exponential family and as such, can be modelled in the framework of GLMs. Here, the marginal distribution of continues covariates is assumed to be Gaussian. This may be a strong assumption for the covariates used in the insurance applications. Many of them are defined on the positive domain and as such, require to be modelled with a non-Gaussian distribution (e.g. density or exposure).

In our TCWM approach, we further augment Equation 1.1.3 to accommodate modelling of continues covariates through a transformation function $\phi(.)$.

$$f(\boldsymbol{x}, y; \Phi) = \sum_{g=1}^{G} \tau_g f(y|\boldsymbol{x}, \vartheta_g) \phi\{f^T(\boldsymbol{v}, \theta_g)\} f(\boldsymbol{w}, \theta_g) \tag{1.1.4}$$

where $f^T(\boldsymbol{v}, \theta'_g)$ represents the transformed function of $f(\boldsymbol{v}, \theta'_g)$. The $\phi(.)$ function can be applied to any continuous covariates that are non-Gaussian. The existence of $\phi(.)$ is only possible if there exists a bijective map from a non-Gaussian, to a Gaussian. If we consider the case where not all continuous covariates need transformation, we can further split the continues covariates $\boldsymbol{V} = (\boldsymbol{U}', \boldsymbol{T}')'$ to those that need transformation $\boldsymbol{U}$ and those that do not need transformation $\boldsymbol{T}$. Due to independence, Equation 1.1.4 can be easily modified as follows

$$f(\boldsymbol{x}, y; \Phi) = \sum_{g=1}^{G} \tau_g f(y|\boldsymbol{x}, \vartheta_g) \phi\{f^T(\boldsymbol{u}, \theta''_g)\} f(\boldsymbol{t}, \theta''_g) f(\boldsymbol{w}, \theta'''_g). \tag{1.1.5}$$

We denote Equation as the 1.1.5 transformed generalized linear mixed (TGWM) model. Here, the term "transformed" allows us to relax the Gaussian assumption for some continues covariates by introducing a transformation to Gaussian distribution, the term "generalized linear" refers to the conditional distribution of $Y|\boldsymbol{x}$, the term "mixed" emphasizes the presence of both continues and discrete (mixed type) covariates.

1.1.1 Transformation Function Selection

The choice of transformation functions depend on the existence of a bijective map from a non-Gaussian to Gaussian distribution. In this paper, we focus on the log-normal distribution as it is relevant to most actuarial applications. We aim to show the existence of a relevant transformation function beginning with the univariate log-normal distribution denoted as $\mathcal{LN}(x; \mu, \sigma)$.

$$\mathcal{LN}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$$

We then use the property that a change of variables must conserve differential probability.

$$\mathcal{LN}(x; \mu, \sigma) = \frac{\mathcal{N}(\ln x; \mu, \sigma)}{x} dx = \mathcal{N}(\ln x; \mu, \sigma)\frac{d\ln x}{dx} dx = \mathcal{N}(\ln x; \mu, \sigma) d\ln x$$

Thus log normal distribution can be written as a Gaussian form.

$$\mathcal{N}(\ln x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right] = \mathcal{LN}(x; \mu, \sigma) \times x \tag{1.1.6}$$

We then extend this identity to the multivariate log-normal case where $f^T(\boldsymbol{u}; \mu'_g, \Sigma'_g)$ is the multivariate log-normal distribution with component vectors $\theta'_g = (\mu'_g, \Sigma'_g)$.

$$f^T(\boldsymbol{u}; \mu'_g, \Sigma'_g) = \frac{1}{(\prod_{i=1}^{N} u_i)|\Sigma_g|(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\ln(u) - \mu_g)'\Sigma_g^{-1}(\ln(u) - \mu_g)\right]$$

By the identity (2.6), $f^T(\boldsymbol{u}; \mu'_g, \Sigma'_g)$ is transformed into a Gaussian distribution.

$$f(\ln \boldsymbol{u}; \mu'_g, \Sigma'_g) = f^T(\boldsymbol{u}; \mu'_g, \Sigma'_g) \times (\prod_{i=1}^{N} u_i)$$

6

As a result the relevant transformation function for a log-normal to Gaussian distribution for some arbitrary covariate x.

$$\phi(f^T(x;\theta_g)) = f^T(x;\theta_g) * (\prod_{i=1}^{N} x_i)$$

. With the relevant transformation function shown to exist the TGWM model is rewritten as:

$$f(\boldsymbol{x}, y; \Phi) = \sum_{g=1}^{G} \tau_g f(y|\boldsymbol{x}, \vartheta_g) f(\ln \boldsymbol{u}, \theta_g') f(\boldsymbol{t}, \theta_g'') f(\boldsymbol{w}, \theta_g''). \tag{1.1.7}$$

1.1.2 Parameter Estimation

In most finite mixture problems, the standard method for estimating the optimal number of components G is based on the expectation-maximization (EM) algorithm proposed by **?** and further discussed by **?**. The EM algorithm is based on the maximum likelihood estimation. The initial values of the parameter estimates are generated through a stochastic initialization, then the algorithm proceeds by alternation of the E-step and M-step in order to update the parameter estimates as well as missing data which in this case correspond to the posterior probability that x_i comes from the gth mixture component, computed at each iteration of the EM algorithm. In order to find an optimal number of components, maximum likelihood estimation is obtained over a range of G, and the best model is selected based on a chosen model selection criterion. the convergence of the EM algorithm is achieved when the relative increase in the log-likelihood function is no bigger than a small pre-specified tolerance value or the number of iterations reach a limit. In this subsection, we explain the parameter estimation in line with the CWM methodology proposed by **?**.

Proposed TGWM model is based on the assumption that $f(y|\boldsymbol{x}, \vartheta_g)$ belongs to the exponential family of distributions that are strictly related to GLMs. The link function in Equation 1.1.1 relates the expected value $g(\mu_g) = \beta_{0g} + \beta_{1g}x_1, \ldots, +\beta_{pg}x_p$. We are interested in estimation of the vector $\boldsymbol{\beta}_g$, thus the distribution of $\boldsymbol{y}|\boldsymbol{x}, \mathcal{R}_g$ is denoted by $f(\boldsymbol{y}|\boldsymbol{x}, \beta_g, \lambda_g)$, where λ_g denotes an additional parameter to account for when a distribution belong to a two-parameter exponential family.

The marginal distribution $f(\boldsymbol{x}, \theta_g)$ has the following components: $f^T(\boldsymbol{u}, \theta_g')$, $f(\boldsymbol{t}, \theta_g'')$, and $f(\boldsymbol{w}, \theta_g''')$. The first two components are modelled as Gaussian density with mean $\boldsymbol{\mu_g}$ and covariance matrix $\boldsymbol{\Sigma}_g$ as $\phi(\boldsymbol{v}, \boldsymbol{\mu_g}, \boldsymbol{\Sigma}_g)$.

The marginal density $f(\boldsymbol{w}, \theta_g''')$ assume that each finite discrete covariate W is represented as a vector $\boldsymbol{w}^r = (\boldsymbol{w}^{r1}, \ldots, \boldsymbol{w}^{rc_r})'$ where $w^{rs} = 1$ is w_r has the value s, s.t. $s \in \{1, \ldots, c_r\}$, $w^{rs} = 0$ otherwise.

$$f(\boldsymbol{w}, \boldsymbol{\gamma_g}) = \prod_{r=1}^{q} \prod_{s=1}^{c_r} (\gamma_{grs})^{w^{rs}}, g = 1, \ldots, G \tag{1.1.8}$$

where $\boldsymbol{\gamma_g} = (\gamma_{g1}', \ldots, \gamma_{gq}')'$, $\boldsymbol{\gamma_{gr}} = (\gamma_{gr1}', \ldots, \gamma_{grc_q}')'$, $\gamma_{grs} > 0$, and $\sum_{s=1}^{c_r} \gamma_{grs}, r = 1, \ldots, q$. The density $f(\boldsymbol{w}, \boldsymbol{\gamma_g})$ represents the product of q conditionally independent multinomial distributions with parameters $\boldsymbol{\gamma_{gr}}, r = 1, \ldots, q$.

Let $(\boldsymbol{x}_i, y_i), \ldots, (\boldsymbol{x}_n, y_n)$ be a sample of n independent observations drawn from model in (1.1.4). For this sample, the complete data likelihood function, $L(\boldsymbol{\Phi})$, is given by

$$\text{Ł}_c(\boldsymbol{\Phi}) = \prod_{i=1}^{n} \prod_{g=1}^{G} \left[\tau_g f(y_i|x_i, \boldsymbol{\beta}_g, \lambda_g) \phi\{f^T(u_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)\} f(t_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) f(w_i, \gamma_g) \right]^{z_{ig}} \qquad (1.1.9)$$

where Z_{ig} is an indicator variable with value of $Z_{ig} = 1$ indicating that observation $(\boldsymbol{x_i}, y_i)$, originated from the g-th mixture component and $z_{ig} = 0$ otherwise.

By taking the logarithm of (1.1.9), the complete data log-likelihood function, $\ell_c(\boldsymbol{\Phi})$, is written by

$$\ell_c(\boldsymbol{\Phi}) = \sum_{i=1}^{n} \sum_{g=1}^{G} z_{ig} \left[\log(\tau_g) + \log f(y_i|x_i, \boldsymbol{\beta}_g, \lambda_g) + \log \phi\{f^T(u_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)\} + \log f(t_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + \log f(w_i, \gamma_g) \right]$$

$$(1.1.10)$$

It follows that at the sth iteration, the conditional expectation of Equation 1.1.10 on the observed data and the estimates from the $(s-1)$th iteration results in

$$Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(s-1)}) = \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \left[\log(\tau_g) + \log f(y_i|x_i, \boldsymbol{\beta}_g, \lambda_g) + \log \phi\{f^T(u_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)\} \right.$$

$$\left. + \log f(t_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + \log f(w_i, \gamma_g) \right]$$

$$= \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \log(\tau_g) + \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \log f(y_i|x_i, \boldsymbol{\beta}_g, \lambda_g) + \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \log f(t_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$$

$$+ \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \log \phi\{f^T(u_i, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)\} + \sum_{i=1}^{n} \sum_{g=1}^{G} \tau_{ig}^{(s-1)} \log f(w_i, \gamma_g)$$

The idea behind the EM algorithm is to generate a sequence of the estimates from the maximum likelihood estimation starting from an initial solution $\hat{\boldsymbol{\Phi}}^1$ and iterating it with the following steps until convergence. Here, the E-step computes $Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(s)})$ and M-step computes $\Phi^{(s)} = \arg\max_{\Phi \in \mathscr{R}} Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(s)})$. The E-step does not depend on the form of density, and the latent data only relate to $\boldsymbol{z}$. The posterior probability that $(\boldsymbol{x_i}, y_i)$ comes from the gth mixture component is calculated at the sth iteration of the EM algorithm as

$$\tau_{ig}^{(s)} = E[Z_{ig}|(\boldsymbol{x_i}, y_i), \boldsymbol{\Phi}^{(s)}] = \frac{\tau_g^{(s)} f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta_g}^{(s)}, \lambda_g^{(s)}) \phi\{f^T(\boldsymbol{u}_i, \boldsymbol{\mu}_g^{(s)}, \boldsymbol{\Sigma}_g^{(s)})\} f(\boldsymbol{t}_i, \boldsymbol{\mu}_g^{(s)}, \boldsymbol{\Sigma}_g^{(s)}) f(\boldsymbol{w}_i, \gamma_g^{(s)})}{f(\boldsymbol{x_i}, y_y; \boldsymbol{\Phi}^{(s)})}.$$

$$(1.1.11)$$

The denominator of (1.1.11) corresponds (1.1.4).

The M-step requires maximization of the Q-function with respect to $\boldsymbol{\Phi}$ which can be done separately for each term on the right hand side in Equation 2.11. As a results, the parameter estimates $\hat{\tau}_g$, $\hat{\boldsymbol{\mu}}_g$, $\hat{\boldsymbol{\sigma}}_g$, and $\hat{\boldsymbol{\gamma}}_g$, are obtained on the $(s+1)$th iteration as follows

$$\hat{\tau}_g^{(s+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau_{ig}^{(s)},$$

$$\hat{\boldsymbol{\mu}}_g^{(s+1)} = \frac{1}{\sum_{i=1}^{n} \tau_{ig}^{(s)}} \sum_{i=1}^{n} \tau_{ig}^{(s)} \ln \boldsymbol{u}_i,$$

$$\hat{\boldsymbol{\sigma}}_g^{(s+1)} = \frac{1}{\sum_{i=1}^{n} \tau_{ig}^{(s)}} \sum_{i=1}^{n} \tau_{ig}^{(s)} (\ln \boldsymbol{u}_i - \hat{\boldsymbol{\mu}}_g^{(s+1)})(\ln \boldsymbol{u}_i - \hat{\boldsymbol{\mu}}_g^{(s+1)})',$$

$$\hat{\gamma}_{gr}^{(s+1)} = \frac{\sum_{i=1}^{n} \tau_{ig}^{(s)} \omega_i^{rs}}{\sum_{i=1}^{n} \tau_{ig}^{(s)}},$$

while the estimates of $\boldsymbol{\beta}$ are computed by maximizing each of the G terms

$$\sum_{i=1}^{n} \tau_{ig}^{(s)} \log f(y_i | \boldsymbol{x}_i, \boldsymbol{\beta}_g, \lambda_g) \tag{1.1.12}$$

Maximization of Equation 1.1.12 is performed by numerical optimization in R software in a similar framework the mixture of generalized linear models are implemented. For additional details about this implementation the reader is refer to **?** and **?**. For insurance applications, current TCWM model can be used for modelling frequency of claims assuming that $\boldsymbol{Y}$ belongs to Poisson or Binomial distributions. When modelling severity of claims, $\boldsymbol{Y}$ can be assumed accommodate Gamma or Lognormal distributions. All of these applications are based on CWM as the underlying approach. For additional information, the reader is referred to the "**flexCWM**" manual for **R** users written by Mazza A., Punzo A., and Ingrassia S. (2015).

1.1.3 Model Selection Criterion

Model performance is assessed by computing the values of Akaike Information Criterion (**?**), and Bayesian Information Criterion (**?**), using the following formulas

$$AIC = 2k + 2LL$$
$$BIC = k\ln(n) + 2LL$$

where k represents the number of estimated parameters in the model and n is the sample size. Lower values of AIC and BIC indicate better model fit to the data. While we compute all of these measures, the final model selection decision is based on the BIC criterion. **?** suggested that BIC performs better in mixture modelling and as such it is proffered over AIC. To be consistent with the most of the literature in this area, we will make the final decision on model selection based on the BIC criterion.